DOI: 10.1007/s11676-007-0049-1

Estimation of forest parameters based on TM imagery and statistical analysis

CHEN Wen-bo[1*], ZHAO Xiao-fan[2,3]

[1] *College of Land Resources and Environment, Jiangxi Agricultural University, Nan Chang 330045, P.R. China*
[2] *Institute of Applied Ecology, Chinese Academy of Science, Shenyang 110016, P.R. China*
[3] *Graduate School of Chinese Academy of Science, Beijing 100039, P.R. China*

Abstract: One of the primary forestry research interests lies in estimating forest stand parameters by applying empirical or semi-empirical model to establish the relationship between the forest stand parameters and remote sensing data. Using remote sensing image and the inventory data from 2 compartments in northeast Florida, U.S.A., this paper explored the correlation between forest stand parameters and Landsat TM spectral digital number (DN) value. Results showed that less than 50% of the total variance could be explained by linear regression models with only either a single band or such vegetation indices as vegetation index (VI) or normalized difference vegetation index (NDVI) as predicators. In consequence, multi-linear regression models which synthesized more predicators were introduced to estimate forest parameters. Regression results were tested in terms of the other group of data, and verification showed a better capability of explaining over 75% variance except for forest density. The weakness and further improvement of prediction models were also discussed in the article. This paper is expected to provide a better understanding of the relationship between TM spectral and forest characteristics

Keywords: TM image; DN value; Estimation of forest parameters; Correlation and regression analysis

Introduction

Recently remote sensing has been intensively used in forestry and revolutionized the methodology of traditional inventory. Forest stand parameters, such as age, volume or biomass, average stem diameter and height, are important to assess forest resources (Mallinis *et al.* 2004). Traditional inventory of forest stand parameters based on fieldwork is often difficult, costly and time-consuming to conduct, especially in a large area (Lu *et al.* 2004). TM image, as a widely used remote sensing data, has long time been used as an auxiliary data to obtain information on forest resources (Reese *et al.* 2002; Makela and Pekkarinen 2004). These auxiliary data, together with field inventory data can thus be used for the estimation of forest characteristics.

It is essential to understand how image data relate to forest stand characteristics (Lucas and Curran 1999; Visser and Stampfer 2003). Empirical models including correlation and regression are important tools for relating field-measured stand parameters to remote sensing data. The strength of them is that they describe the measured data using a specified mathematical function or curve (Peng *et al.* 2002). In addition, regression analysis is easier to link two sets of data for providing continuous estimation of stand variables (Khatry Chhetri and Fowler 1996; Zas *et al.* 2004). Hence, regression models are popularly applied due to their plainness and easy to be constructed and understood.

The main objective of this study is to estimate forest parameters on the basis of TM data. Sub-objectives are to (1) study the correlation between individual band of TM data and forest parameters selected, such as forest age, forest density, mean basal area per tree and mean diameter at breast height per tree; (2) explore the correlation between the most commonly used vegetation indices (NDVI and VI) and forest parameters selected and (3) evaluate the relationship between different band combination and forest parameters by regression model and determine which band or band composition is the most suitable for corresponding stand parameters and why by stepwise regression.

Materials and method

Site description

The study area is located in Duval County (30°19' 24" N, 81° 40' 29" W) of northeast Florida (Fig. 1). Physiographically, it is within Coastal Plain, which is characterized by a series of ancient marine terraces oriented parallel to the present coastline. The area has relatively flat terrain and is less than 100 m in elevation. Sandy soil is predominant soil type in the study area and according to the climatological data it has mean annual precipitation of 1,352 mm and mean annual temperature of 20.4ºC. Most of the study area is forested, primarily with two southern pine species, slash pine *(Pinus elliottii)* and longleaf pine *(Pinus palustris)*. Another major forest type is bottomland swamp forests occurring on areas of relatively poor drainage. Pond cypress *(Taxodium distichum var. nutans)*, bald cypress *(Taxodium distichum)* and other broadleaf hardwoods are the dominant species occurring in these swamp forests.

Received: 2007-05-09; Accepted: 2007-06-26

Electronic supplementary material is available in the online version of this article at http://dxdoi.org/10.1007/s11676-007-0049-1
Biography: CHEN Wen-bo (1974-), male, Ph.D., College of Land Resources and Environment, Jiangxi Agricultural University, Nan Chang 330045, P.R. China. Email: cwb1974@126.com
Responsible editor: Chai Ruihai

Data collection

Inventory Data

Field data were used to construct relationships between the measured forest parameters and TM data. During the fieldwork, the forest stand boundary maps and Landsat TM color composites were used to allocate the samples sites, and GPS devices were used to identify the coordinates of each site.

In this study, 60 stands (30 m×30 m for each) were randomly selected within 2 different compartments, excluding the stands near or cross the boundary of the image data. 40 stands data were used for regression model construction and the other 20 stands were used for validation. The species of the stands were identified as slash pine, longleaf pine, pond cypress and bald cypress, occupying 37%, 29%, 21% and 13%, respectively. The forest stand parameters selected for this study were average age (AGE, yr), forest density expressed by trees per square meter (FD, tree·m^{-2}), mean basal area per tree (MBA, m^2·tree^{-1}), mean diameter at breast height per tree (MDBH, m·tree^{-1}).

Image data

A cloud-free frame of Landsat-5 TM data, collected on Oct. 12, 1984 (Fig. 1) was selected in this study. To ensure compatibility between image and the ground data, radiometric and atmospheric correction and geometrical rectification of remotely sensed data were performed before the research. An improved image-based dark object subtraction (DOS) model was used to correct atmospheric effect in the study. Geometric coordinate was transformed and registered from imagery to a standard local map projection to facilitate linkage with the ground data in the presence of a set of ground control points (GCPs) defining a mathematical transformation equation. Nearest neighbor resampling was applied in geometrical transformation in order to minimize changes of the statistical properties of the data sets. In the images obtained from the TM scanner, each pixel was represented by a digital number DN from 0 to 255. Zonal statistics of GIS and correlation & regression of statistics were applied in this study.

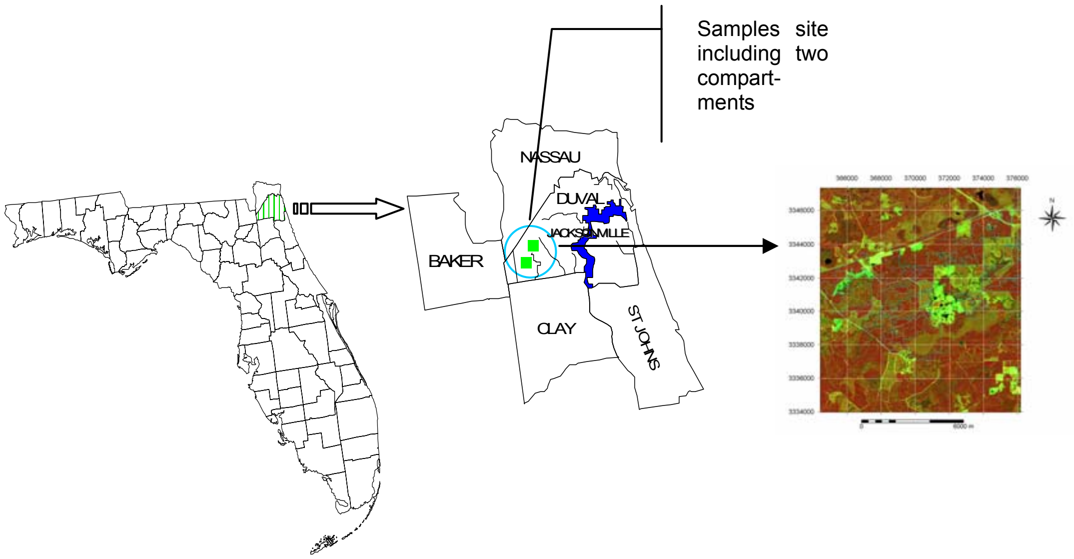

Fig. 1 Location of samples site and forest cover TM image (4 band in red, 5 in green and 2 in blue) of study area

Results and discussion

Correlation between forest parameters and TM image

A good understanding of the relationships between forest stand parameters and remote sensing spectral was a prerequisite for effectively using appropriate image bands to estimate forest stand parameter. The average of TM image digital number (DN) values and four forest parameters selected in each of 40 stands were applied to conduct correlation analysis. Coefficients of Pearson's correlation p and of determination R^2 were calculated. Correlation with p-value less than 0.05 was assumed to be statistically significant. Moreover, R^2 was also used to predict the power of description of the regression model. The results were shown as table 1.

Although most of bands statistically had a statistically significant linear relationship with almost all the parameters selected, it did not mean that the linear regression models established were good enough for parameters estimation. Since most of absolute r were less than 0.7, only less than 50% of the variance can be explained by the regression model.

Correlation between forest parameters and vegetation index

Relationship between vegetation indices and forest parameters were explored in this section. As it is well known, vegetation strongly reflects infrared electromagnetic wave and absorbs red one. Thus the difference in vegetation absorbing infrared and red can be considered as a crucial indicator and the important basis of various vegetation indices. Numerous indices were constructed before. In this section, two most commonly applied

vegetation indices NDVIN (normalized vegetation index with a power of n) and VI (vegetation index) were used. NDIVN was calculated as (NIR^n – Red^n)/ (NIR^n + Red^n) (n = 1, 2, 3). The results were shown as Table 1.

Table 1 also showed that for VI or NDVIN, all R^2 were less than 50%, presumably half of the variance was able to be explained by the linear regression models. Thus, vegetation index as predicator only, the same as individual band can not be made a convincible prediction model.

Table 1 Correlation between TM band DN value and forest stand parameters (n=40)

Pearson correlation		TM1	TM2	TM3	TM4	TM5	TM7	VI	NDVI1	NDVI2	NDVI3
AGE	coefficient	-0.378*	-0.258	-0.449*	0.122	-0.640*	-0.593*	0.583*	0.564*	0.543*	0.508*
	p	0.006	0.065	0.001	0.388	0.000	0.000	0.000	0.000	0.000	0.000
	R^2	0.142	0.067	0.202	0.015	0.410	0.352	0.340	0.318	0.295	0.258
FD	coefficient	-0.283*	-0.341*	-0.276*	-0.367*	-0.205	-0.167	0.162	0.184	0.197	0.214
	p	0.042	0.013	0.047	0.007	0.144	0.236	0.250	0.193	0.161	0.127
	R^2	0.080	0.116	0.076	0.135	0.042	0.028	0.026	0.034	0.039	0.046
MDBH	coefficient	-0.463*	-0.383*	-0.542*	0.09	-0.738*	-0.675*	0.702*	0.671*	0.642*	0.595*
	p	0.001	0.005	0.000	0.524	0.000	0.000	0.000	0.000	0.000	0.000
	R^2	0.214	0.147	0.294	0.008	0.545	0.456	0.493	0.450	0.412	0.354
MBA	coefficient	-0.406*	-0.325*	-0.485*	0.121	-0.675*	-0.623*	0.643*	0.614*	0.586*	0.542*
	p	0.003	0.019	0.000	0.392	0.000	0.000	0.000	0.000	0.000	0.000
	R^2	0.165	0.106	0.235	0.015	0.456	0.388	0.413	0.377	0.343	0.294

Note: * means significance at 0.05 level (two-tailed t-test)

Regression between forest parameters and TM image

Neither single band nor vegetation index was capable of predicting forest parameters. In order to establish a model that can explain more variance, more explanatory variables were required to be involved. Consequently, stepwise multi-linear regression method was applied in this study. Multiple stepwise regression analysis generated a linear equation that predicted a dependent variable as a function of several independent ones. Construction of the polynomial was performed by a best-fitting selection of the polynomial terms using the stepwise regression algorithm. The number of terms for satisfying a user-given accuracy of approximation was minimized. For each coefficient, the t-ratio tests whether the value of the coefficient was zero or not, and if its p-value was less than 0.05, the calculated value was considered statistically significant. Table 2 illustrated the results.

Table 2 Coefficients of stepwise multi-linear regression models for forest stand parameters

Model	coefficient	Unstandardized Coefficients β	Standardized Coefficients β^*	Standardized Error	t	p	R^2
AGE	constant	-405.061		105.115	-3.853	0.000	0.738
	TM2	15.141	2.564	1.992	7.600	0.000	
	TM3	-12.578	-4.114	1.590	-7.913	0.000	
	TM1	4.405	1.193	2.164	2.036	0.047	
FD	constant	1583.657		514.317	3.079	0.003	0.135
	TM4	-22.720	-0.367	8.139	-2.791	0.007	
MDBH	constant	-36.894		7.163	-5.151	0.000	0.759
	TM5	-0.260	-1.699	0.025	-10.514	0.000	
	TM1	0.727	1.067	0.110	6.604	0.000	
MBA	constant	-2.099		0.443	-4.738	0.000	0.673
	TM5	-0.013	-1.642	0.002	-8.730	0.000	
	TM1	0.039	1.073	0.007	5.705	0.000	

For the parameter of age, the best linear regression model was established using TM band 1, 2 and 3 as predicators. This model explained 73.8% of total variance (R^2) and band 3 was the most deterministic predicator due to the t-ratio value (-7.913). For FD, only one predicator band 4 was left after stepwise regression. Although the model was statistically significant, it could only explain 13.5% of total variance and was poor for prediction. For MDBH, the best linear regression model was established using TM band 1 and 5. This model explained 75.9% of total variance, of which band 5 had comparatively stronger influence according to the t-ratio value. For MBA, the best linear regression model was established using band 1 and 5. The multi-linear model explained 67.3% of total variance and band 5, as it was of MDBH, had a relatively stronger influence.

By means of stepwise regression, the regression models established showed a stronger ability to explain more variance compared to that with simple band or vegetation indices. The regression models listed provide the choices for predicting forest stand parameters.

$$\mathrm{AGE} = -405.061 + 15.141 * TM2 - 12.578 * TM3 + 4.405 * TM1$$

$$\mathrm{FD} = 1583.657 - 22.720 * TM4$$

$$\mathrm{MDBH} = -36.894 - 0.260 * TM5 + 0.727 * TM1$$

$$\mathrm{MBA} = -2.099 - 0.0133 * TM5 + 0.0389 * TM1$$

The other 20 stands data were used to verify the prediction results. The results of verification showed the prediction precision of 79%, 34%, 81% and 75% for AGE, FD, MDBH and MBA respectively.

Conclusion

Forest stand characteristics can be described by many stand parameters such as age, DBH, density and height, but these data collection are time and labor consuming by conventional forest inventory. Remote sensing is normally considered an alternative and assumed to be a better method for obtaining insight of forest characteristics. TM image data and linear regression model are commonly used to fulfill this task.

The relationship between forest parameters selected and TM image DN values was explored by linear regression model in the article. It is concluded that the regression models with only single band or vegetation index as predicator are not reliable for applying prediction. Multi-linear regression model which synthesizes multi-band of TM as dependant variables is necessary to be introduced. The case study showed that except for FD, all other forest parameters can be explained and predicted by stepwise regression models, offering a better understanding of relationships between bands of TM image and forest characteristics

A weakness of this study is that the sample set was not equally weighted due to varied area of forest stand, which will more or less reduce the credibility of prediction models. Nevertheless, the general tendency of the interaction relationship between TM data and forest parameters is not changed. Moreover, the regression model for prediction could not be transferred to other areas because the data were not calibrated to their true energy level. In order to obtain a transferable model, the calibration from DN value to reflectance should be performed firstly, which tremendously strengthens the flexibility of prediction model over the time and space.

References

Khatry Chhetri, D.B., Fowler, G.W. 1996. Estimating diameter at breast height and basal diameter of trees from stump measurements in Nepal's lower temperate broad-leaved forests. *For. Ecol. Manage.*, **81**(1): 75–84.

Lu, D., Mausel, P., Brondizio, E., *et al*. 2004. Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *For. Ecol. Manage*, **198**(1-3): 149–167.

Lucas, N.S., Curran, P.J. 1999. Forest ecosystem simulation modelling: the role of remote sensing. *Prog. Phys. Geog*, **23**(3): 391–423.

Makela, H., Pekkarinen, A. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *For. Ecol. Manage.*, **196**(3): 245–255.

Mallinis, G., Koutsias, N., Makras, A., *et al*. 2004. Forest parameters estimation in a European Mediterranean landscape using remotely sensed data. *For. Sci*, **50**(4): 450–460.

Peng C.H., Liu J.X., Dang Q.L., *et al*. 2002. TRIPLEX: a generic hybrid model for predicting forest growth and carbon and nitrogen dynamics. *Ecol. Mode.,* **153**: 109–130.

Reese, H., Nilsson, M., Sandstrom, P., *et al*. 2002. Applications using estimates of forest parameters derived from satellite and forest inventory data. *Comput. Electron. Agr.*, **37**(1): 37–55.

Visser, R., Stampfer, K. 2003. Tree-length system evaluation of second thinning in a loblolly pine plantation. *South. J. Appl. For.*, **27**(2): 77–82.

Zas, R., Merlo, E., Diaz, R., *et al*. 2004. Relative growth trend as an early selection parameter in a Douglas-fir provenance test. *For. Sci*, **50**(4): 518–526.